Pedagogic research in anatomical sciences: a best practice guide

Article (Accepted Version)

Cover Page

Pedagogic Research in Anatomical Sciences: A Best Practice Guide

C.F. Smith[1*], R. I. Stabile[3], G.M. Finn[2]

[1]Department of Medical Education, Brighton and Sussex Medical School, University of Sussex, Brighton, United Kingdom

[2]Hull York Medical School, University of York, York, United Kingdom

[3]International School for Foundation Studies, University of Malta

Running title: Research in Anatomy

*Correspondence to: Dr. Claire F. Smith, Brighton and Sussex Medical School, University of Sussex, Medical School Building, Falmer, BN1 9PX, United Kingdom. E-mail: c.smith@bsms.ac.uk

SUMMARY

This article explores the background of anatomical educational research. It draws together research and our own personal experiences to propose a best practice piece for novice researchers in anatomical education. The article explores the domains of both qualitative, and quantitative methods as applied to anatomy pedagogy. It takes into consideration validity and what might be undertaken to increase validity and reliability. The article explores how both qualitative and quantitative data can be analysed and recommends top tips including: Identify your research questions and theoretical framework. Map out how you are going to answer your research questions. Consider collaborating with like-minded researchers in other countries: multi-centre studies have a better chance of getting published and carefully consider your target journal and suggestions for peer review, taking into consideration individuals expertise and potential conflicts of interests. This article is designed to be a guide to anyone starting anatomical research or experienced researchers looking for new methods and ideas.

Key Words: learning anatomy, qualitative, quantitative, pedagogical research, focus groups, questionnaires

INTRODUCTION

The landscape of anatomical pedagogy has developed considerably within the last ten years (Drake et al., 2009). Early research papers were concerned with the 'how to teach' and the prosection and dissection debate, although this continues (Smith, 2008). To begin with, many educational papers often reported what they did, while providing little evidence or theoretical base for their research. Over time, the quality and quantity of anatomical pedagogical research articles has increased, especially in peer-reviewed journals. Within anatomy education there has been a long-standing problem of educators having to defend themselves. It may be questions such as 'How much anatomy should we teach?', 'What do students feel about this experience?', 'Are the assessments testing re-call or understanding?' These types of questions are common to all anatomy educators because not only do we care but we should also defend what we do and why. As anatomists began to engage more in these questions, the interest in anatomy education has grown.

The emergence of 'Anatomical Sciences Education' recognised the rising role of anatomical pedagogy and gave anatomists and educators a dedicated output for dissemination. At the same time, there was a drive to align anatomical educational research with the same quality and rigor that is applied to other educational research. Around this time, Computer Assisted Learning (CAL) was novel and innovative. Today, Technology Enhanced Learning (TEL) is less of a subcategory and often completely integrated into everyday anatomical teaching. The move to understand more of the student experience has involved a range of different facets of educational theory including: learning approaches (Smith and Mathias, 2007; Ward, 2011), personality (Finn et al., 2015), assessment (Smith and McManus, 2015), near peer teaching (Hall et al., 2013), 3D

printing (Li et al., 2015; Lim et al., 2016; Smith et al., 2017) and inter-professional learning (McBride and Drake 2015; Smith et al., 2015) to name a few.

These aspects of educational theory all started with a common goal: to understand a component of learning anatomy, they asked a question. This is the first step in any research. It is important to design a high-quality research question. Anatomy education research has stemmed from different facets of a broad range of research areas including educational theory and psychology. This article draws on a mixture of these backgrounds to provide an evidence base for undertaking quality anatomy pedagogical research. The purpose of this article is to help establish some of the fundamentals that a new researcher (or one who is experienced yet new to pedagogical research) might need to know. We also seek to offer a best practice approach to continue to increase the validity of the discipline.

DESIGNING RESEARCH QUESTIONS

Designing a high-quality research question is the first step in any research endeavour. It is important to design research questions that precisely focus on the phenomena of interest. It is common to start with a broader question such as, how do students learn anatomy? This main question can then be sub-divided to guide the research activities. For example:

- What are medical students' perceptions of anatomy?
- How are medical students approaching anatomy learning?
- What is involved in the learning process?
- What affects and influences the learning of anatomy?
- How are medical students applying their anatomy knowledge?

- How is anatomy linked to clinical practice?

In devising research questions, one approach is to write out the question and sub-questions several times and ask a colleague to discuss them. A good question to ask yourself is the three 'W's' (why, what and when). Why does this matter? What would the research impact be? When can this research be undertaken and how? Having designed good quality questions, it is good practice to assess what components may affect the study and when these might occur. A time plan or Gantt chart is essential to map out the study. This should include details such as the time frame for ethical review, when data gathering will occur, data analysis and dissemination plans.

ETHICAL CONSIDERATIONS

Ethical concerns related to the study need to be considered and must adhere to local guidelines. In anatomy, the subject being explored might be a sensitive issue for students. The relationship between stakeholder and researcher needs to be considered e.g. if a lecturer is asking for sensitive information or is asking a question that may divulge sensitive information that the student has not previously disclosed.

Informed consent should be key and all participants should know that their involvement is voluntary. Participants should beforehand receive a thorough explanation of the benefits, rights, risks, and dangers involved because of their participation in the research project (Frankfort-Nachmias & Nachmias, 1992). Participants who have given consent also have the right to refuse to take part or to withdraw at any point in time without prejudice to the participant. It is standard to give all potential participants a copy of the consent form and Participant Information Sheet (PIS). This may not be possible in some projects, for example, where a questionnaire is asking

participants to respond. Here a consent statement explaining that by participating in the survey they are giving consent, can be used instead. Depending on the nature of the study it may be necessary to treat participants with anonymity and confidentiality. To anonymise participants a coding system can be established, preferably by a third party. However, if participants agree to face-to-face interviews, total anonymity would obviously not be possible. In this case, all information received should be treated as confidential. Confidentiality can be achieved by a coding strategy with confidential data being securely stored under password control and, if necessary, appropriately disposed of. All activities for the study should be considered under Health and Safety regulations and, where appropriate, regulations such as the Human Tissue Act of 2004. A sample consent form and participant information sheet is provided in Figure 1. Should an ethics panel not grant approval, formal feedback can be helpful to work through the reasons. For example, it may be that the researcher would be asking questions in a focus group about how they perceive intimidation within a viva, but the researcher concerned often takes part in viva examinations, and hence the ethical panel would have considered a conflict of interest which could affect the integrity of the study. In this example, the study may be permitted if a trained researcher in this methodology could be found from another department. Further information on ethics as applied to anatomical research can be found at http://www.ifaa.net/committees/ethics-and-medical-humanities-ficem/.

THE METHODOLOGICAL STANCE

As a main principle, there is a divide between qualitative and quantitative research. In higher education, we generally intend to encourage the development of conceptual understanding in students, so a method which so vividly portrays differing conceptualisations must have direct

relevance to learning and teaching (Entwistle, 2000). Several methodologies might need to be considered to investigate research questions. For example, longitudinal follow up of a group of students or a cross-sectional study of a cohort, or possibly a randomised controlled trial comparing one intervention with another.

In designing a research study, it is useful to map out the study design (Figure 2). Such a map will also prove helpful to any ethics committee and will ensure alignment between different groups of participants or different phases of a study. Naturally a study may be a simple one-population-one-methodology case study or it may involve different groups and different methodologies as shown in Figure 2.

SAMPLING

In designing an anatomical educational research project, the sample group of participants needs to be considered. A sample is defined as a portion or subset of a larger group called a population (Fink, 2003), with the best sample being representative of the population characteristics. The main advantages of selecting a sample is that it saves time as well as financial and human resources; however, the disadvantage is that one estimates or predicts information rather than establishing the information from a whole population. (Kumar, 1996). There are two types of sampling: probability (random) and non-probability (purposive) (Cohen, Manion & Morrison 2001).

In any educational research design, the aim should be to study individuals who are representative of the general population of interest such that the results can be generalisable. Otherwise the

external validity of the study is said to be low. One way to achieve this is through random sampling (which is different from random allocation), which samples units from a defined population such that each participating unit has the same chance of being selected. This can be very difficult to attain because randomly selected units (individuals, student cohorts, universities etc.) may decline to participate. Therefore, frequently, educational research focuses on so called convenience samples which are easier to recruit, monitor and follow-up, resulting in generally good response and retention rates. Nevertheless, volunteers may differ from non-volunteers e.g., in the case of anatomy educational research, they may perceive themselves as weaker and hence choose to participate in a study of an intervention designed to improve outcomes. External validity is also reduced is there is a high non-response rate, as it may well be that those who choose not to respond are somehow different from those who do, thus affecting the generalisability of the results. Ideally, study participants who for whatever reason fail to complete the whole intervention should be included in the analysis to avoid systematic errors. Depending on the research question, the methodology will guide the sample size. In the case of a qualitative study it may be suitable to interview only four participants. On the other end of the scale it may be suitable for a qualitative study to survey a cohort of 500 students.

QUALITATIVE METHODS

Qualitative research is primarily exploratory. Wrongly, it is often regarded as subordinate to quantitative research. Qualitative research methodologies can be utilised to gain an understanding of experiences, opinions and motivations. They are often employed to provide insight into a research question or to develop a research hypothesis for quantitative studies.

Qualitative methods can provide meaning and insight into why the results of a quantitative investigation are so, or because the researcher needs to understand the local context. Qualitative research can take many formats ranging from interviews and focus groups to analysis of imagery or observational studies. Within anatomy education, interviews and focus groups are frequently employed methods of data collection as they are useful methods for exploring perceptions and experiences.

Interviews

Interviews involve one participant and can assume multiple formats: structured, semi-structured or unstructured. These terms refer to the nature of questioning and exploration within the interview and could be considered on a continuum. A structured interview is rigid; it presents each participant with the same questions in the same order. Semi-structured interviews involve the use of some pre-defined questions and prompts but allow for the interviewer to be responsive and probe in response to a participant's answers. This enables comparisons to be made between participants in a group whilst allowing for unexpected issues to also be freely discussed. This creates a fluid and flexible format (Mason, 2002). At the other end, an unstructured interview does not use any pre-defined questions and is often driven by the participant. An example of an interview study is that of Collier et al., (2012) who conducted semi-structured interviews with graduate teaching assistants to evaluate the introduction of technology in the classroom with a focus on student performance and student evaluations.

Focus Groups

Data collection from groups is known as focus group discussions or focus group interviews.

Focus groups typically consist of between four and eight participants. The purpose of a focus group is to listen and gather information to understand how people feel and think about an issue (Krueger, 2000). Focus group discussions are guided but unstructured in questioning, facilitated by a moderator or researcher. A focus group interview utilises a list of questions, similar to a semi-structured or structured interview, but has numerous participants. An example of a focus group study is that of Finn and McLachlan (2010) who explored students' views on the use of body painting as a learning tool.

Observation

Observation may be selected to enable a better understanding of a case (Stake, 1995). Using observation, a researcher removes her/himself from the teaching environment of being a teacher and, either through watching live or from video recordings, aims to understand phenomena such as the behaviours of a group. This perspective is referred to as Ethnography (Clifford and Marcus, 1986) which is the scientific study of people and cultures together with their customs, habits, and mutual differences. Semi-structured observations can capture elements but also still allow for unexpected aspects to be observed (Cohen, Manion & Morrison 2001). To record an observation event sampling may be used, when a line or tick is made against an activity (e.g. student opened book). Instantaneous sampling may also be used, where the events are recorded in chronological order. In addition to the structured recording, handwritten note observations can be made. Structured observations and field notes can be analysed through categorical indexing (Mason, 2002) or coding to produce an overview of the activities and interactions in the setting.

Phenomenology and Phenomenography

Another type of qualitative research is phenomenological research, the goal of which is to describe a lived experience of a phenomenon. It involves the in-depth analysis of narrative data focussing on the meaning of the experience, behaviour and narrative of participants. In essence, a phenomenological research study explores answers to the question 'What is it like to experience…?'. Researchers explore multiple perspectives on the same situation to establish some generalisations of what the experience is like from an insider's perspective.

In a similar vein, there is also phenomenography. Phenomenography and phenomenology are not the same thing. Phenomenography is empirical and investigates the experiences of others. Phenomenology is interpretive and explores the phenomenon itself, whereas the focus of phenomenography explores the experiences and the subsequent perceptions of the phenomenon (Cibangu and Hepworth, 2016).

Analysis

Data from interviews and focus groups are transcribed, usually verbatim, before being analysed. Data analysis is often completed in iterative cycles, whereby one interview or focus group is held, transcribed and analysed before conducting the next. Transcripts can be coded using free nodes for content (words) and meaning. Figure 3 shows a sample transcript where the coding has been added in by hand. This can also be undertaken online utilising dedicated software such as NVivo (NVivo qualitative data analysis Software; QSR International Pty Ltd). Commonly utilised methods of data analysis include thematic analysis, discourse analysis and grounded theory, although the latter is both a methodology and a paradigm. Thematic analysis is defined by Braun and Clarke (2006) as: "A method for identifying, analysing and reporting patterns within data." (p. 79). Thematic analysis is a widely-used method of analysis in qualitative

research and enables the researcher to generate simple themes within their dataset. Discourse analysis (Wodak and Meyer, 2009) focuses on talk, text and images. It critiques the way that topics have been conceptualised and treated. A specific example of a discourse analysis is metaphor analysis when the metaphors utilised by participants are studied to provide meaning (Schmitt, 2005). Grounded theory was developed by Glaser and Strauss (1999) – it is regarded as both a methodology and paradigm (Finn and McLachlan, 2010). Grounded theory utilises open, axial and selective coding to produce a theory that is 'grounded' within the data. It is further characterised by iterative cycles of data collection with constant comparison and member checking (Lingard and Kennedy, 2007).

Validity and Reliability in Qualitative Research

Validity is the extent to which the findings of a study can be applied to other situations (Vanderstoep and Johnston, 2009), in other words, the generalisability. It is due to this perceived lack of generalisability that qualitative research is sometimes dismissed as being less worthy. Generalising the findings of a qualitative study is difficult, as the sampling is often purposive, and the findings largely contextual. The validity of a qualitative study must therefore come from the transferability of the findings (Merriam, 2009). It has been suggested that the best way to achieve transferability is to therefore paint as full a picture as possible of the context and findings of the study. Another suggested method for improving transferability is careful selection of the study sample. Maximum variation is a sampling method (Merriam, 2009), which will allow for the application of the study to a wider audience. This variation is achievable by increasing the

sites used for a study, choosing more participants or selecting a typical sample of participants (Merriam, 2009).

Reliability, with respect to qualitative methodologies is problematic. This is principally because human behaviour is not static, nor is one opinion more likely to be correct over another. In quantitative terms, reliability is achieved when a study can be repeated yielding the same results (Vanderstoep and Johnston, 2009). For qualitative research, replication may never be achieved. This does not discredit the data as it should be acknowledged that data can yield many interpretations. Considering this, the reliability of qualitative data comes from asking whether the results are consistent with the data collected (Merriam, 2009).

Measures of reliability, such as Cronbach's alpha and Kendall's tau are sometimes utilised on qualitative data, particularly from surveys or questionnaires (Roland and Cooil, 1994). The primary aim of qualitative research is not to achieve a measure of statistical significance.

QUANTITATIVE METHODS

Quantitative research is the systematic gathering of data that are analysed via statistical analysis to prove/disprove a theory. Quantitative methods often include surveys and randomised controlled trials.

Surveys/Questionnaires

Quantitative research in anatomy education most often utilises surveys that are either cross-sectional (at one point in time) or longitudinal (at several points in time). The former is most appropriate for collecting descriptive information, whereas the latter, if appropriately timed, may be useful in addressing analytic cause-and-effect questions. The group of individuals being studied is known as the cohort (e.g., of students, of graduates etc.), and hence such surveys are also known as cohort studies. In the case of longitudinal studies, a change in the magnitude of the measured variable of interest (e.g., exam pass marks, etc.) between baseline and follow-up time points is compared with that of the comparison group, although this is not always possible.

Surveys are often used as they enable a large amount of quantitative data to be gathered, although some questionnaires may also contain qualitative aspects. Survey questions are designed to fit into a scale (frequently a Likert scale) that provides a range of suitable responses. Survey design must ensure that questions flow and are sequential (Cohen, Manion & Morrison 2001; Foddy, 2001). Piloting the questions helps to ensure that they ask what they are supposed to. Potential drawbacks of surveys should also be considered. For example, high non-response rates might affect the validity of the method (Cohen, Manion & Morrison 2001). However, triangulation from other methods e.g. focus groups can ensure that the results obtained are reasonably representative.

Survey response rates are increased and bias is decreased by simple, well-designed questions that are easily understood, as the types of responses received are easily influenced by the wording, the form and the order of questions. The layout (now often online using tools such as Survey Monkey or Google Forms) should have clear instructions, be easy to read, and avoid distractors

such as unnecessary colours. The questions themselves should be simply worded and care should be taken to avoid double negatives. Questions containing two questions are ambiguous and should not be used. Leading or loaded questions also introduce bias, resulting in criticism that surveys can be constructed to show exactly what the investigator wishes.

Open-ended, free response questions are better suited when the subject is complex, or the possible replies are too many to pre-code, or possibly not all known, but these data are more difficult to analyse. However, closed questions may force respondents to choose among possibly less appropriate categories. They may also clue-in respondents to options they may not have thought of themselves. Pilot testing is important to ensure your questions are asking what you want them to.

The form of the survey is also important (Bradburn and Sudman, 1974). For example, respondents are more likely to agree with a statement than disagree with its opposite, which is why standardised questionnaires vary the response categories as well as the direction of the question (Goldberg and Williams, 1988). The likelihood of stereotyped responses is increased when a series of questions is asked with similar response formats. For example, if the scales are always the same, respondents are more likely to agree with a statement on the right-hand side of the scale. Thus, questionnaires should alter the direction of the response codes and avoid using the same response formats repeatedly. This helps the respondent think about the question rather than reply automatically. Response formats can be dichotomous (Yes/No), multiple choice, or scaled (Always, Mostly, Sometimes, Rarely, Never).

In general, surveys should start with easier, non-threatening questions, followed, unless there are reasons not to, by the most important questions (so that important information is not lost if the respondent fails to complete it). General questions on the same topic should precede specific ones, as the latter can influence the former, and questions about attitudes should be asked after questions about behaviour (Bourque and Fielder, 1995).

Apart from dichotomous Yes/No formats, the most commonly used scaling method for measuring responses is the Likert format, because it is easily understood, analysed and interpreted (Likert, 1932). Other attitudinal scaling methods in use are the Thurstone (1928) and Guttman (1944) techniques. The former has been widely used in studies of the attitudes of medical students to anatomy (Moxham et al,. 2016). Each of these methods assumes that a numerical score can be used to represent a person's attitude towards the question being asked.

The Likert scale usually encompasses five points scale e.g., Strongly Agree, Agree, Undecided, Disagree, Strongly Disagree (coded from 5 to 1). However, respondents who prefer not to think about the question at all, will generally opt for "undecided" as it is simply easier to answer. This leads some researchers to avoid offering the middle option altogether by presenting a four or six-point scale. Analysis of data by simply adding scores assumes that all items are of equal importance to all respondents, which is unlikely. Similarly, assuming equal intervals between each score can introduce bias. Statistical tools such as factor analysis can be used to calculate appropriate weighting to each item score.

Depending on the geographical location of the study, it may be necessary to consider translation

of validated tools and their cultural equivalence. Standardised surveys are generally reflective of cultural norms (Guyatt, 1993), which is why simple translation and back-translation (into the original language) may not be sufficient to ensure congruence between words and their true meaning in the translated language. White and Elander (1992) have described a methodological approach to this issue, which is becoming more important as multi-centre studies flourish.

Statistical analysis is beyond the scope of this paper. However, for the sake of completeness, descriptive or univariate statistics is used to analyse one variable, while bivariate tools are used to analyse the association between two variables e.g., correlations. Multivariate analyses allow one to measure the effect of one variable on the outcome measure while controlling for the effects of other variables e.g., multiple or logistical regression (Campbell and Swinscow, 2010).

Controlled Trials

A more rigorous research method is the experimental approach, wherein the educational intervention (e.g., flipped classroom, peer teaching, spatial ability training etc.) also known as the independent variable, is systematically manipulated by the investigator under controlled conditions with an equivalent comparison group. This approach significantly reduces the possibility of bias (or systematic errors) and chance (or random errors). Ideally, participants would be randomised (Moser and Kalton, 1971) to either experimental or control groups, and ideally, there should be a pre-test of both groups that precedes the intervention, and a post-test of both groups thereafter. Random allocation to experimental or control groups means that participants have an equal chance of being assigned to either group, thus minimising the risk of confounding the results by an extraneous variable. Although this research design can yield causal

relationships, it may be expensive, difficult to set up, and sometimes unnatural in that it can be difficult to control the experimental environment. Also, educational research participants may be reluctant to consent to random allocation to interventions. Moreover, randomisation does not exclude the possibility that the randomised population, e.g., anatomy students using 3D printed models in Australia, may not be typical of the world-wide population of anatomy students (Lim et al, 2015).

Quasi-experimental studies may also incur bias by the so-called Hawthorne effect which occurs when awareness of being investigated results in study participants changing their behaviour (Parsons, 1974). This is only one of the many sources of bias in any experimental design, some of which can be minimised by "blinding" the study participants such that they do not know whether they have been randomised to the intervention or control group. Clearly this is very difficult to achieve when examining the effect of an educational intervention.

In real-life educational settings, creating an experimental design to investigate a research question is not always possible. Although less ideal, several other methods can be used e.g., pre- and post-testing without a control group, a so-called observational study. For example, Zhang et al., (2017) showed that the average Pre/Post anatomy MCQ score increased by 39% after participating in six hour-long doughnut round sessions. In this case, statistical tools of co-variance adjustment can be used to decrease bias.

Studies using non-randomised control groups (with matched controls e.g., for age, gender, year

of study, etc. who have not been exposed to the intervention) are much easier to conduct. For example, Stoner et al., (2017) found no significant difference in exam marks between students who did/did not watch online gastro-intestinal anatomy videos in a flipped classroom study. However, even if the intervention and control groups are very similar, the reliability of the conclusions will always be limited by the non-randomised design. Replicating the study findings in several settings can improve the reliability of the results. In the absence of randomisation, statistical tools (e.g., cross tabulation, standardisation and regression) can be used to control for extraneous variables that are known.

Another quantitative research approach might be a time-series method, where for example different student cohorts are studied over different phases of the intervention period (so-called historical controls). Trends in data over time can then be compared with data collected from other sources.

Individuals could also act as their own controls, and the study could collect data about them before and after the intervention. This type of study design can be influenced by other events unrelated to the study, but could still be useful to generate hypotheses. However, in the absence of a comparison non-randomised control group, it can never really be known whether the study findings could have occurred anyway without the intervention being studied.

In summary, all non-randomised experimental methods have the potential for bias e.g., students in the control group who unknown to the investigator, are exposed to an intervention which is very similar to the one being investigated (e.g., Iqbal et al, 2017). Thus, any observed differences

cannot be unequivocally ascribed to the intervention.

A cross-over study is a longitudinal study in which participants receive a sequence of different exposures. Cross-over designs are popular for education research as students receive all interventions (and/ or the control) and therefore are not missing out on learning opportunities, making them favourable from an ethical and educational perspective. An example of this is the study of Finn et al., (2010) who explored the impact of an authentic context (wearing hospital scrubs) on learning and recall.

Mixed Method

The validity of any research approach is enhanced by triangulation i.e., the use of three or more methods. For example, a survey questionnaire may be complemented by face-to-face interviews and/or focus groups and/or observation, as was shown in Figure 2. Bias is reduced when the same variables are measured by research methods with different methodological weaknesses. Furthermore, triangulation of data which measures the same variables at different times, places or groups may also serve to overcome the limitations of one research method and/or of a particular investigator. Hence the call for replication studies (Artino, 2013). It is very important that all research subjects be accounted for in the analysis i.e., all individuals who withdraw or drop-out of the study should be included in the analysis. The so-called "healthy survivor effect" when applied to health research, applies also to attrition from *any* longitudinal study as it may result in artificially improved follow up results.

VALIDITY AND RELIABILITY

It is important that the researcher acknowledges that they are part of the world in which they are researching and therefore cannot be completely objective. Cohen et al. (2001) explain that validity in its earliest form was based on the view that it was essential that the measuring device measured what it was supposed to. Validity provides certainty and confidence in the results. These include the appropriateness of the overall methodological framework, literature searching, the sampling strategies and methods adopted. Triangulation can be an important way to increase validity. Triangulation can also be through investigator triangulation (Stake 1995) especially in elements which are open to greater subjectivity. For example, analysis of data collected via focus groups could be performed by another colleague to ensure that the results are a true presentation of the data and that the same conclusions are drawn.

Reliability refers to the consistency of the results. It is possible to improve reliability through equivalence and internal consistency. Equivalence may be enhanced through inter-rater reliability (Cohen, Manion & Morrison 2001). Internal consistency judges the reliability of the instruments used (Trochim, 2006). In quantitative research this is achieved by performing Cronbach's Alpha tests that measure the reliability of a scale, and whether the items used are measuring the same thing and are comparable (Bland & Altman, 1997, Boone and Boone, 2012). Each data collection method has its own advantages and disadvantages as well as potential for bias. Interview bias for example, may "lead" subjects if the order of pre-coded response choices might mean subjects to answer in a particular way.

A question often asked is "how many responses do I need?" Although there is no generally

agreed standard minimum acceptable response rate, non-response rates less than 25% are considered "acceptable". Adding a covering letter from a legitimate person/institution, incentives (not necessarily financial ones), and regular reminders may enhance the response rate. Item non-response is said to occur when study subjects choose not to answer *some* of the questions. The longer the structured questionnaire is, the more likely it is that subjects will omit questions, especially the more difficult/important ones which are often incorrectly placed at the end. Hence, the importance of considering your questionnaire design to ensure it encourages participation. There are statistical methods of handling randomly missing data which are beyond the scope of this article.

DISSEMINATION

Dissemination of results should be planned at the start of the project to determine the more suitable place for the topic area. Dissemination may take the form of conference presentations, manuscripts in journals, seminars, online blogs and engagement in a wide range of social media and other networks.

The key to dissemination is to firstly create a catchy title e.g., the main findings of the study (Iqbal et al, 2017) and a clear abstract, both of which will attract attention. When submitting an abstract to a conference, care should be taken to submit work that fits well within the theme and remit of the conference. Some conferences publish peer-reviewed abstracts as proceedings and the abstracts gain a Digital Object Identifier (DOI) number which is used to permanently identify an article.

The ultimate goal is publication of the manuscript resulting from the research study in a peer-reviewed journal with a high impact factor. Journals take a wide range of articles, and it is important to choose your first-choice journal and the type of article (original research, review article, letter, commentary etc.) very carefully. It may be suitable to split a large study into several articles or it may be necessary to join smaller pieces of research together to make up a full article. In anatomical education, it is important to ensure that the context of the education is described in a manuscript as this can vary between institutions and countries.

CONCLUSIONS

Anatomical pedagogical research is an important and growing area which has gained status as it continues to improve educational outcomes for students, patients and the wider public.
Such research should answer a question of interest or usefulness and should always be founded on ethically sound principles. Anatomical research may involve a range of qualitative and quantitative methodologies. All research should be able to demonstrate impact in different domains which may be economic and/or social. The findings of the research should always be fed back into the local curriculum, and where possible made generalisable and transferable to be of benefit to the wider anatomical community.

TOP TIPS FOR GETTING STARTED

- Identify your research questions and theoretical framework. Map out how you are going to answer your research questions.

- Conduct a literature review to determine whether your research question has already been examined. Do not forget to look at theses and dissertations as these often provide useful follow-up ideas for further research
- Consider collaborating with like-minded researchers in other countries: multi-centre studies have a better chance of getting published.
- Consider a small-scale pilot project to scope out the area of interest.
- Start an early dialogue with your local ethics committee.
- Establish roles within the research team.
- Clearly organise your methods e.g. select participant groups, create incentives for participants e.g. snacks or vouchers.
- Don't worry it you hit a rate limiting step e.g. lack of participation; understand why and make suitable adjustments (within ethics approval).
- Ask a colleague to check through your data interpretation.
- Write your work up with colleagues keeping clear track of changes and versions.
- Carefully consider your target journal and suggestions for peer review, taking into consideration individuals' expertise and potential conflicts of interests.

ACKNOWLEDGEMENTS

The authors wish to thank all of those involved in anatomy education, especially the Education Committee of the Anatomical Society and the Trans-European Pedagogic Anatomical Research Group (TEPARG) for ongoing discussions that inform anatomical education theory and practice.

REFERENCES

BLAND JM, ALTMAN DG (1997) Cronbach's Alpha. *BMJ*, 22:314.

STAKE RE (1995) The Art of Case Study Research: Sage Publications, Inc.

ARTINO AR (2013) Why don't we conduct replication studies in medical education? *Med Educ*, 47:746-447.

BOONE HN, BOONE DA (2012). Analysing likert data. *Journal of extension* 50.2:1-5.

BOURQUE LB, FIELDER EP (1995) How to conduct self administered and mail surveys. Thousand Oaks, CA: Sage Publications, Inc.

BRADBURN NM, SUDMAN S (1974) Improving interview method and questionnaire design. San Francisco, CA: Jossey-Bass.

BRAUN V, CLARKE V (2006) Using thematic analysis in psychology. Qualitative Research in Psychology, 3 (2). pp. 77-101. ISSN 1478-0887 Available from: http://eprints.uwe.ac.uk/11735

CAMPBELL MJ AND SWINSCOW T DV (2010) Statistics at Square One, 11th Edition, BMJ Books

CIBANGU SK AND HEPWORTH M (2016). The uses of phenomenology and phenomenography: A critical review. *Library & Information Science Research* 38.2:148-160.

CLIFFORD J, MARCUS GE (1986) Writing Culture: The Poetics and Politics of Ethnography, University of California Press.

COHEN L, MANION L, MORRISON K (2001) *Research Method in Education*, 5th edn,

COLLIER L, DUNHAM S, BRAUN MW, O'LOUGHLIN VD (2012) Optical versus virtual: Teaching assistant perceptions of the use of virtual microscopy in an undergraduate human anatomy course. *Anat Educ Sci*, 5:10-19.

DRAKE RL, MCBRIDE JM, LACHMAN N, PAWLINA W (2009) Medical education in the anatomical sciences: The winds of change continue to blow. *Anat Sci Educ*, 2:253–259.

Entwistle N (2000) 'Promoting deep learning through teaching and assessment: conceptual frameworks and educational contexts', *Teaching Learning Research Programme conference.* Leicester.

FINK A (2003) *How to sample in surveys,* Sage.

FINN GM, MCLACHLAN J (2010) A qualitative study of student responses to body painting. *Anat Sci Educ,* 3:33-38.

FINN GM, PATTEN D, MCLACHLAN JC (2010) The impact of wearing scrubs on contextual learning. *Medical Teach*, 3*2*:381-384.

FINN GM, WALKER SJ, CARTER M, COX DR, HEWITSON R, SMITH CF (2015) Exploring relationships between personality and anatomy performance. *Anat Sci Educ*, 8:547–554.

FODDY W (2001) *Constructing questions for interviews and questionnaires* Cambridge, University Press.

FRANKFORT-NACHMIAS C, NACHMIAS D (1992), *Research Methods in the Social Sciences,* London: Edward Arnold.

GLASER G, STRAUSS A (1999) *The Discovery of Grounded Theory: strategies for qualitative research,* Aldine Transaction.

GOLDBERG D, WILLIAMS P (1988). A users guide to the general health questionnaire. Windsor: NFER-Nelson.

GUTTMAN L(1944) A basis for scaling quantitative data. *Am.Sociol.Rev,* 9:139-50.

GUYATT GH (1993) The philosophy of health-related quality of life translation. *Quality of Life Research,* 2:461-5.

HALL S, STEPHENS J, ANDRADE T, DAVIDS J, POWELL M, BORDER S (2013) Perceptions of junior doctors and undergraduate medical students as anatomy teachers: Investigating distance along the near-peer teaching spectrum. *Anat Sci Educ,* 7:242–247.

IQBAL H, GALEA M, AGIUS A, STABILE I (2017) Clinical N-PAL Preferentially Benefits International Students. Med EdPublish **DOI:** https://doi.org/10.15694/mep.2017.000186

KRUEGER R (2000) *Focus groups. A practical guide for applied research*, 3rd edn, Sage.

KUMAR R (1996) *Research Methodology. A step-by-step guide for beginners,* Addison Wesley Longman Australia Pty Limited.

LI Z, LI Z, XU R, LI M, LI J, LIU Y, SUI D, ZHANG W, CHEN Z (2015) Three-dimensional printing models improve understanding of spinal fracture–A randomised controlled study in China. *Sci Rep,* 5:11570.

LIKERT R (1932) A technique for the measurement of attitudes. *Archives of Psychology*, 22:1-55.

LIM KH, LOO ZY, GOLDIE SJ, ADAMS JW, MCMENAMIN PG (2016) Use of 3D printed models in medical education: A randomised control trail comparing 3D prints versus cadaveric materials for learning external cardiac anatomy. *Anat Sci Educ,* 9:213–221.

LINGARD L, KENNEDY TJ (2007) Grounded Theory. *Qualitative research in medical education.* Edinburgh: Assocation for the Study of Medical Education (ASME).

MASON J (2002) *Qualitative researching*, 2nd edn, Sage.

MCBRIDE JM, DRAKE RL (2015) Student perceptions of an interprofessional educational experience: The important of goal articulation. Anat Sci Educ, 8:381-385.

MERRIAM S. *Qualitative Research: A Guide to Design and Implementation*. San Francisco, CA Jossey-Bass; 2009.

MOSER CA, KALTON G (1971) Survey methods in social investigation, 2nd edition. London: Heinemann.

MOXHAM BJ, EMMANOUIL-NIKOLOUSSI E, STANDLEY H, BRENNER E, PLAISANT O, BRICHOVA B, PAIS D, STABILE I, BORG J, CHIRCULESCU A (2016) The Attitudes of Medical Students in Europe towards the Clinical Importance of Embryology. *Clin Anat,* 29:144–150.

PARSONS HM (1974) What happened at Hawthorne? *Science* 183:922-32.

RUST RT, COOIL R (1994) Realiability Measures for Qualitative Data: Theory and Implications. *Journal of Marketing Research, 31:1-14.*

SCHMITT R (2005) Systematic metaphor analysis as a method of qualitative research. *The qualitative report,* 10:358-394.

SMITH CF, TOLLEMACHE N, COVILL D, JOHNSTON M (2017) Take away body parts! an investigation into the use of 3D-printed anatomical models in undergraduate anatomy education. Anat Sci Educ **DOI:** 10.1002/ase.1718.

SMITH CF, MATHIAS H (2007) An investigation into medical students' approaches to anatomy learning in a systems-based prosection course. *Clin Anat*, 20:843–848.

SMITH CF, MCMANUS B (2015) The integrated anatomy practical paper: A robust assessment method for anatomy education today. *Anat Sci Educ,* 8:63–73.

SMITH CF, HALL S, BORDER S, ADDS PJ, FINN GM (2015) Interprofessional anatomy education in the United Kingdom and Ireland: Perspectives from students and teachers. *Anat Sci Educ,* 8: 360-370

SMITH CF (2008) Learning anatomy at university: effectiveness, issues and implications for the future education of doctors (Doctoral thesis) University of Southampton. https://ethos.bl.uk/Logon.do;jsessionid=C2982A98941BF6CCCF4F321378ACF204?ordering=1

STONER R, CARUANA C, STABILE I (2017). Head over Heels in Gastrointestinal Anatomy: A Case study. *MedEdPublish* **DOI:** https://doi.org/10.15694/mep.2017.000107

THURSTONE LL (1928) Attitudes can be measured. *American Journal of Sociology*, 33: 529-54.

TROCHIM WM (2006) The Research Methods Knowledge Base, 2nd Edition. Internet WWW page, at URL: <http://www.socialresearchmethods.net/kb/>

VANDERSTOEP S, JOHNSTON D. *Research methods for everyday life: blending qualitative and quantitative approaches*. San Francisco, CA Josey-Bass; 2009

WARD PJ (2011) First year medical students' approaches to study and their outcomes in a gross anatomy course. *Clin Anat,* 24:120–127.

WHITE M, ELANDER G (1992) Translation of an instrument. The US Nordic family dynamics nursing research project. *Scandinavian Journal of caring Science* 6:161-4.

WODAK R, MEYER M. (Eds.). (2009). Methods for critical discourse analysis. Sage.

ZHANG Y, ZERAFA- SIMLER MA, STABILE I (2017) Supported self-directed learning of Clinical Anatomy: a pilot study of doughnut rounds. *Eur J Anat,* 4:319-324.